

UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF TEXAS
AUSTIN DIVISION

Voxer, Inc. and Voxer IP LLC,

Plaintiffs,

v.

Facebook, Inc. and Instagram LLC,

Defendants.

Civil Action No. 1:20-cv-00655-ADA

Jury Trial Demanded

**DECLARATION OF DR. MICHAEL MITZENMACHER IN SUPPORT OF VOXER'S
RESPONSE TO DEFENDANTS' MOTION FOR JUDGMENT ON THE PLEADINGS**

I, Michael Mitzenmacher, declare and state as follows:

I. INTRODUCTION

1. I am over the age of 18 and am competent to make this Declaration. I have personal knowledge or have developed knowledge of these technologies based upon my education, training, or experience of the matters set forth herein.

2. I have been retained by counsel for plaintiffs Voxer, Inc. and Voxer IP LLC (collectively, "Voxer") as an expert to provide opinions regarding questions of fact relevant to the patent eligibility of the asserted claims of U.S. Patent Nos. 8,180,030 (the "030 patent") and 10,142,270 (the "270 patent") (collectively, the "asserted patents").

3. I am being compensated at my usual hourly rate of \$900 per hour. I am also being reimbursed for any out-of-pocket expenses. I have no financial interest in, or affiliation with, Voxer. My compensation does not depend in any way on the outcome of this case, the substance of my testimony, or opinions that I express.

4. In rendering my opinions, I have considered the patents at issue and their file histories; the parties' patentable subject matter briefs, the parties' claim construction disclosures,

briefs, and expert declarations to date; the Court's claim construction order; my prior claim construction declarations; and any other documents referenced, discussed, or listed in this declaration. I have also utilized my own knowledge and experience in computer networking, data structures, algorithms, multimedia processing, and other computing fields relevant to the asserted patents. I have also reviewed the productions associated with the claim construction disclosures in this case, including the extrinsic and intrinsic evidence cited by the parties.

5. My analysis of the materials produced in this matter is ongoing, and I will continue to review any new material as it is provided. This declaration represents only those opinions I have formed to date. I reserve the right to amend or supplement my opinions based on additional documents or evidence I am presented, including without limitation any arguments or expert declarations advanced by Facebook in this case.

II. QUALIFICATIONS AND EXPERIENCE

6. I hereby incorporate the qualifications I discussed in my Declaration in Support of Voxer's Opening Claim Construction Brief, including my Curriculum Vitae attached thereto as Exhibit A. Dkt. 61-1 ¶¶ 8-14.

III. LEGAL STANDARDS

7. I understand that determining patent eligibility requires a two-step process.

8. In step one, I understand that a court must decide if the claims are directed to an abstract idea. I understand that this may entail deciding whether the claims are focused on improving computer technology or if they merely use computer technology as a tool to implement an abstract idea.

9. In step two, I understand there is a question of fact whether the claims include an "inventive concept," something more than well-understood, routine, and conventional activities in the industry, sufficient to confer patent eligibility.

IV. TECHNICAL BACKGROUND

10. As the '270 and '030 patents both note, existing communication systems allowed a user to engage in “full-duplex, synchronous conversation” or to transmit “one-way asynchronous voice message[s],” but there was a divide between those two options. '030 patent at 1:37-57.¹

11. Of relevance to at least the '030 patent, existing communication systems in common use lacked a “communication device” with persistent storage of incoming and outgoing streams, and only allowed the recipient to listen or view the resultant stream of data in real time (in the case of a synchronous conversation) or time-shifted (in the case of an asynchronous message). The recipient in these existing communication systems lacked an option to render a received streaming media in both a real-time mode and a time-shifted mode.

12. Of relevance to at least the '270 patent, another limitation of such existing communication systems is that they required that a connection be first established before any voice or video data is transmitted. '270 patent at 1:41-44 (“One is still required to pick up the phone, dial another party, ***wait for a connection to be made***, and then engage in full-duplex, synchronous conversation with the dialed party.”) (emphasis added); *id.* at 1:63-67 (“Yet another problem with current voice mail systems is that ***a connection must be made*** between the caller and the voice mail system ***before a message can be left***. If no connection is made, there is no way for the caller to leave a message.”) (emphasis added); *id.* at 2:42-48 (“Most tactical radio communication must occur through a ‘live’ radio connection between the sender of a message and a recipient. If there is no radio connection between the two parties, there can be no

¹ The '030 and '270 patents share the same specification. I am citing to just the '030 patent specification for convenience, but the same disclosure is also found in the '270 patent.

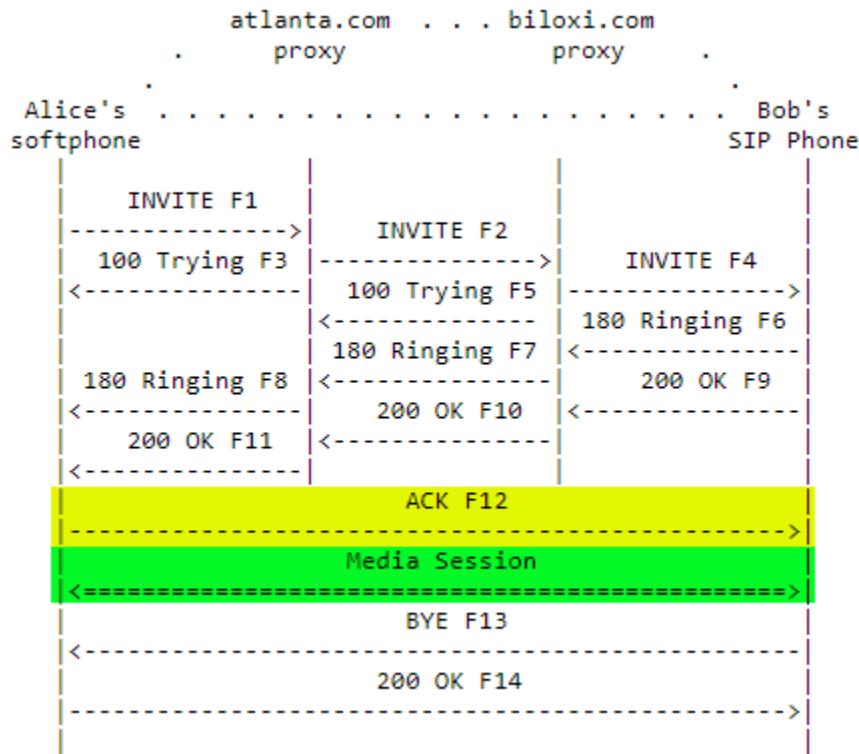
communication. Urgent messages cannot be sent if either the sender or the receiver does not have access to their radio, or a radio circuit connection is not established.”).

13. I previously addressed the term “end-to-end connection” in my Declaration in Support of Voxer’s Opening Claim Construction Brief (Dkt. 61-1 ¶¶ 34-40) and my Declaration in Support of Voxer’s Responsive Claim Construction Brief (Dkt. 62-1 ¶¶ 12-17). The “ends” refer to source or destination devices in a network, in contrast with other devices that may lie on the network path between the source and the destination. For there to be an “end-to-end connection,” the endpoints would need to be able to directly communicate with one another (possibly via the other network devices) by, for example, addressing packets to the other endpoint’s IP address.

14. While these concepts can be understood even by a person without specialized training, that does not render them any less technical. An end-to-end connection is a concrete, technical property of a communication system that has real effects on system performance. For example, waiting to establish a connection before commencing data transmission introduces latency and thereby decreases performance of a communication system. The latency required to establish a connection may not mean much in the context of asynchronous communication, such as e-mail, where a few seconds make little difference, but this latency becomes significant in the context of real-time communication. Moreover, requiring an end-to-end connection before commencing transmission decreases the system’s flexibility in routing the relevant transmission to recipients that may be unknown, unknowable, or lacking a network connection at the start of the transmission, but that may nonetheless subsequently desire and be able to receive that transmission.

15. These limitations in existing communication systems were present not only in traditional telephone communication (for example, using a public switched telephone network (PSTN)), but persisted “[i]n spite of automated switching, high bandwidth networks and technologies such as satellites, fiber optics, Voice over IP (VoIP), wireless and cellular networks.” ’270 patent at 1:37-57.

16. A closer look at VoIP demonstrates how these limitations in existing communication systems continued even in the context of modern communication system. The Session Initial Protocol (SIP) is an example of a conventional VoIP communication system. The June 2002 version of the protocol is available from <https://tools.ietf.org/html/rfc3261> (“RFC3261”). The protocol includes a “SIP session setup example” shown below with annotation:



17. In this example, “Alice’s softphone” is a first endpoint and “Bob’s SIP Phone” is a second endpoint. These two communication devices first go through a signaling stage that

culminates in establishing an end-to-end connection by at least the time “ACK F12” (highlighted in yellow) is transmitted from “Alice’s softphone” to “Bob’s SIP Phone.”

18. Of relevance to at least the ’270 patent, only *after* establishing this end-to-end connection can either communication device begin transmitting real-time voice data as part of a “Media Session” (highlighted in green). The “Media Session” that occurs after the end-to-end connection has been established uses “end-to-end media packets.” RFC3261 at 15.

19. Of relevance to at least the ’030 patent, SIP does not include the capability for a new user to join an existing “Media Session” and then render any part of that conversation in a time-shifted mode. SIP, similar to earlier telephone services, thus limited a “communication device” to rendering an ongoing “Media Session” solely in real-time mode.

20. In the context of the ’030 patent, Facebook’s motion also discusses TV broadcasting, VCRs, and devices for recording telephone conversations. These technologies are not discussed in the ’030 patent, and Facebook’s motion does not provide any evidence regarding specific features or implementations of these technologies. In any event, these technologies lacked a single “communication device” that persistently stored both incoming and outgoing communication and used that capability to render streaming media in both a real-time mode and a time-shifted mode, as recited by the independent claims of the ’030 patent. For example, in Facebook’s analogy to broadcast television and a VCR, Facebook argues the broadcaster can record outgoing streaming media and a VCR can record incoming streaming media, but there is no allegation of a single “communication device” that persistently stores both incoming and outgoing streaming media and provides options to render streaming media in both a real-time mode and a time-shifted mode. Thus, it is my opinion that Facebook’s analogy does

not correspond to the claimed invention of the '030 patent. Facebook's allegations based on recording telephone conversations are similarly lacking.

21. The '030 and '270 patents address these limitations in existing communication systems. Both patents leverage the hybrid real-time/time-shifted architecture described in their specification. Voxer understood that the same messages can be viewed in a real-time mode and also reviewed in a time-shifted mode. To enable the flexibility afforded by this insight, Voxer's architecture (in various embodiments) stored these messages at multiple locations, including the sending device, intermediate server(s), and the receiving device. '030 patent at 4:27-32.

22. The '030 patent claims adds a new capability not previously present in existing communication systems by, for example, enabling a user to "selectively render . . . received streaming media, for the first time, in both a real-time mode as the streaming media is received over the network and in a time-shifted mode by retrieving and rendering the retrieved streaming media from persistent storage." *Id.* at 43:21-26. Existing communication systems lacked this capability: the recipient of a telephone call could participate in that call in real-time, but there was no option to render that call in a time-shifted mode. The recipient of a voicemail could listen to that voicemail later (*i.e.*, in a time-shifted mode), but there was no option to render that voicemail in a real-time mode. The '030 patent claims make this improvement over existing communication systems possible by, for example, having the communication device "progressively and persistently store [the] streaming media received over the network from a remote communication device as the streaming media is received." *Id.* at 4:17-20.

23. The '270 patent claims improve existing communication systems by, for example, allowing a "sending communication device" to transmit real-time "video communication" that is "at least partially rendered . . . while the video communication is transmitted" even "without

having to first establish an end-to-end connection over the network between the sending and the [receiving] communication device.” ’270 patent at 47:37-47. Again, conventional communication systems lacked this capability and required that a connection be established before any real-time transmission could commence. The ’270 patent claims make this improvement over existing communication systems possible by using an intermediate server that “receiv[es] the video communication,” “stor[es] the video communication,” and then “deliver[s] portions of the video communication” to enable the receiving device to begin rendering that video communication. *Id.* at 47:32-42.

V. INVENTIVE CONCEPTS IN THE ASSERTED CLAIMS

24. I have analyzed the claims of the asserted patents, including specifically independent claims 34 and 55 and dependent claims 36-38, 40-41, 43-44, 47-49, and 51 of the ’270 patent, and independent claims 1 and 33 and dependent claims 4, 13, 16, 36, 38, 44, and 47-48 of the ’030 patent.

25. In my opinion, each of these asserted claims includes limitations that, alone or in combination, are more than well-understood, routine, or conventional activity in the industry. I understand that courts refer to such limitations as an “inventive concept.” Below I address exemplary inventive concepts recited in the asserted claims of the ’270 and ’030 patents.

A. Independent Claims 34 and 55 of the ’270 Patent

26. Independent claims 34 and 55 are written from the perspective of an intermediate server that communicates with both a “sending communication device” that creates and transmits “video communication” and a “second communication device” that “at least partially render[s]” the “video communication” while that “video communication is transmitted by the sending communication device.” ’270 patent at 47:23-47; *id.* at 50:3-36. The intermediate server performs the steps of “receiving an identifier associated with a video communication transmitted

by a sending communication device,” “storing the streaming media” for sufficient time to allow the “sending communication device” to begin transmitting the “video communication” even “without having to first establish an end-to-end connection over the network,” “ascertaining, in response to receipt of the identifier, a location on the network for a second communication device,” “receiving the video communication from the sending communication device independently of the location for the second communication device being ascertained,” and “delivering portions of the video communication over the network to the second communication device using the location.” *Id.* By performing this sequence of steps, the intermediate server enables the “sending communication device” to begin transmitting the “video communication” even “without having to first establish an end-to-end connection over the network between the sending and the second communication device.” *Id.*

27. It is my opinion that the '270 patent's particular use of an intermediate server as claimed is contrary to a conventional approach in the field of real-time communication and more than well-understood, routine, conventional activity. For example, in SIP, each endpoint must ascertain the location of all other endpoints because the actual transmission of media occurs using “end-to-end media packets” that generally “take a different path from the SIP signaling messages.” RFC3261 at 15. SIP also does not store these “end-to-end media packets” in the claimed manner at any server used to establish the SIP end-to-end connection because (1) those media packets likely are not routed through the same server; (2) the purpose of SIP is to enable only real-time communication, and thus the creators of SIP did not anticipate a benefit to storing these media packets at an intermediate server or elsewhere; and (3) storing media packets at an intermediate server as claimed was believed to require too much storage and introduce too much latency to be useful for real-time video communication.

28. In addition, the '270 patent allows the “sending communication device” to begin transmitting real-time “video communication” even “without having to first establish an end-to-end connection over the network between the sending and the second communication device.” '270 patent at 47:23-47; *id.* at 50:3-36. This is another claim limitation that is more than “well-understood, routine, conventional activity” in the context of real-time “video communication.” As explained above in the “Technical Background” section, both traditional telephones using PSTN and even VoIP devices communicating over a computer network required *first* establishing an end-to-end connection before either endpoint could begin transmitting voice or other real-time media.

B. Dependent Claims 36-38, 40-41, 43-44, 47-49, and 51 of the '270 Patent

29. Dependent claim 37 requires “start[ing] transmission of the video communication over the network before the location for the second communication devices is ascertained.” *Id.* at 47:61-65. Claim 38 requires “start[ing] transmission of the video communication over the network regardless of whether the second communication device is available or not available on the network.” *Id.* at 47:66-48:4. In my opinion, at least these two steps are more than well-understood, routine, conventional activity in the industry. For example, in SIP, the location of the receiving endpoint had to be ascertained and that receiving endpoint had to be connected to the network before the “Media Session” could take place and the two endpoints could begin exchanging end-to-end media packets. In fact, SIP describes these as two of the “five facets of establishing and terminating media communications: User location: determination of the end system to be used for communication; User availability: determination of the willingness of the called party to engage in communication.” RFC3261 at 8.

30. Dependent claims 41 and 51 require enable a communication device to render video communication “while the second communication is sent by a remote communication

device” and/or “in a time-shifted mode.” ’270 patent at 48:14-23; *id.* at 49:5-15. As discussed below in relation to the ’030 patent, enabling this capability is more than well-understood, routine, conventional activity in the industry.

C. Independent Claims 1 and 33 of the ’030 Patent

31. Independent claims 34 and 55 are directed to a “communication device” that “progressively and persistently store[s] streaming media created using the communication device as the streaming media is created”; “progressively transmit[s] the streaming media over a network as the streaming media is created and persistently stored”; and “progressively and persistently store[s] streaming media received over the network from a remote communication device as the streaming media is received.” ’030 patent at 43:8-26; *id.* at 45:46-64. By performing this sequence of steps, the “communication device” can then “provide rendering options to selectively render the received streaming media, for the first time, in both a real-time mode as the streaming media is received over the network and in a time-shifted mode by retrieving and rendering the retrieved streaming media from persistent storage.” *Id.* at 43:21-26.

32. In my opinion, this sequence of steps is more than well-understood, routine, conventional activity in the industry. As discussed above in the “Technical Background” section, conventional “communication device[s]” were limited to either rendering content in real-time or time-shifted mode, and did not provide the recipient of streaming media “options to selectively render the received media, for the first time, in both” modes. This is at least in part because none of these conventional communication systems performed the same “progressively and persistently stor[ing]” steps recited in the claims of the ’030 patent. Facebook at various times refers to VoIP, broadcast television, and telephonic communication. None of those technologies, even as described by Facebook, included a single “communication device” that was “progressively and persistently storing” both incoming and outgoing “streaming media” or

using such storage to allow “selectively render[ing] the received streaming media, for the first time, in both a real-time mode” and “a time-shifted mode.”

D. Dependent Claims 4, 13, 16, 36, 38, 44, and 47-48 of the '030 Patent

33. Dependent claim 16 requires “transition[ing] the rendering of the received media streams from (i) the real-time mode to the time-shifted mode and (ii) from the time-shifted mode to the real-time mode.” *Id.* at 44:22-26. Dependent claim 36 requires “transitioning communication with the remote communication device between synchronous and asynchronous communication.” *Id.* at 46:8-10. Dependent claim 47 requires “transitioning the rendering of the received media streams from the real time mode to the time-shifted mode.” *Id.* at 46:55-58. And dependent claim 48 requires “transitioning the rendering of the received media streams from the time shifted mode to the real-time mode.” *Id.* at 46:59-62. In my opinion, these are more than well-understood, routine, conventional activity in the industry. In a traditional telephone system, you could not begin listening to a voicemail and then transition that voicemail into a real-time conversation by (for example) switch to listening to the media as it is being received, nor could you pause a real-time conversation and listen to the rest in time-shifted mode. Such capability was also absent from existing VoIP systems, which were similarly limited to just participating in a real-time conversation.

34. Dependent claims 13 and 44 require “(i) ascertain[ing] if the communication device is disconnected from the network when the streaming media is created, wherein disconnected from the network is defined as the inability to transmit the streaming media; and (ii) transmit[ting] the streaming media out of persistent storage when the communication device re-connects to the network, wherein connected is defined as the ability to transmit the streaming media.” *Id.* at 44:4-14; *id.* at 46:37-49. Typical prior art communication systems lacked these capabilities, and these claims are thus more than well-understood, routine, conventional activity

in the industry. For example, in the standard VoIP implementation, the sender had to be connected to a network and establish an end-to-end connection to the recipient before the user could start creating streaming media for transmission. Thus, in VoIP, streaming media was only created as part of an existing “Media Session” and was not transmitted from persistent storage.

35. Dependent claim 38 requires “progressively transmitting indexed media payloads as the streaming media is created, encoded, and persistently stored.” *Id.* at 46:15-18. This is the same “streaming media” that can then be “selectively render[ed] . . . , for the first time, in both a real-time mode” and “a time-shifted mode.” The indexing of “media payloads” is what permits those payload to subsequently be re-assembled into either a “real-time” or “time-shifted” rendering. *Id.* at 23:45-48 (“Since each packet 95 is indexed, time-stamped, and given a sequence identifier, the individual packets can be segmented into Messages. By sequentially threading the individual Messages together, Conversations may be constructed.”). In my opinion, the use of “indexed media payloads” for this purpose is more than well-understood, routine, conventional activity.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on April 19, 2021 in Lexington, MA.



Dr. Michael Mitzenmacher